Validation Test Plan for MIL-STD 2525
Common Warfighting Symbology:
I. Operator Performance Assessment
(13 Sep 96 DRAFT)

I. Objective

The purpose of the validation testing of MIL-STD 2525 warfighting symbology is to evaluate its effectiveness in an operational environment. The validation will include both operator performance testing and an assessment of the symbology in an operational exercise. The test plan presented here addresses the operator testing portion of the validation and describes procedures for assessing both automated and manual rendering of the symbols. This testing will determine if the new symbology provides performance statistics (in terms of speed and accuracy) similar to those for existing symbol sets and if it causes significant confusability problems (e.g., mistaking friend for enemy) for operators.

II. Automated Test Procedure

Automated testing will be performed using existing tactical hardware and operational software and be conducted at various individual and joint service sites (see assumption #1). The software will be instrumented for performance recording so that it can present the test session and record operator responses. The test session will be conducted with the operator seated at normal viewing distance from the workstation monitor. While the focus of the validation will be to assess performance on workstations with high-resolution color monitors and under normal (i.e., standard office) ambient lighting, testing will also include various types and sizes of displays (e.g., low-resolution and monochrome) and other viewing conditions (e.g., bright daylight, low-light, dark-room).

The effectiveness of the new symbology will be assessed by measuring operator performance in a symbol recognition task (see assumption #2). A description of the task (e.g., select all hostile air tracks) will be displayed on the workstation monitor. The operator will click on a Start button to present a set of symbols displayed on a map background and start the clock. The operator will select (i.e., click on) the symbols that match the task and then click on a Done button when finished to stop clock. The elapsed time and number of errors will be recorded, after which the next task description will be presented. The operator will perform symbol selection under time stress to simulate operational conditions and allow the opportunity for errors to occur (so that confusability problems can be identified).

The testing session will begin by presenting a series of symbol recognition tasks using current symbology in order to obtain performance metrics for comparing the data collected on the new symbology. Baseline performance will be measured using the symbol set with which the operator is already familiar (i.e., either "force domain" or "engagement domain" symbology). The operator will then be provided with training on

the new symbology. The training will explain how the new symbols are constructed and provide sufficient practice for the operator to establish a base level of proficiency. The operator will then be given another series of symbol recognition tasks, this time using the new symbology.

The symbology will be presented on a tactical display representative of what the operator would encounter in a joint environment. The STANAG 4420 testing results along with guidelines from the human factors literature will be used as the basis for defining the symbol size, luminance (i.e., symbol/background contrast), color, and font under various viewing conditions. The operator will be tested on symbology at all levels of the symbol hierarchy (i.e., from most complex to most primitive); however, all of the symbols on a given display will be at the same level of the hierarchy. In addition, symbols will be positioned in operationally meaningful groups on each display, with distractor symbols that are similar in appearance to the target(s) included in order to identify potential confusability problems.

The assessment will measure performance with the new symbology when various elements of the tactical display are manipulated. A number of map products will be selected to represent the range of backgrounds upon which the new symbology will likely be displayed. It is expected that these backgrounds will range from single-color (e.g., open ocean displayed in black, gray, or blue) to detailed, multi-color (e.g., terrain elevation data) and include at least five levels of complexity. The presence of tactical graphics will be manipulated to determine the extent to which they impact performance on the symbol recognition task and are confusable with the new symbology. The assessment will measure operator performance when tactical graphics are absent, when they are present to a limited degree, and when they are used extensively. Finally, the symbology will be presented at varying levels of density in order to assess the impact of clutter and overlap among symbols on operator performance.

Testing will be conducted with enlisted and officer operators from each participating organization. Past experience with current symbol sets will be recorded in order to determine if the degree of familiarity with existing symbology has impact on performance with new symbology. Sufficient data will be collected to provide stable assessment of operator performance, and data analysis will include appropriate descriptive and comparative statistics calculated on the each of the performance measures.

III. Manual Test Procedure

Manual testing will be performed in conjunction with the automated testing and limited to operators with experience in this form of symbol rendering. This part of the assessment will measure the extent to which the new symbology can be produced and recognized by operators. In one testing scenario, the operator will be given a template (or shown one of the automated displays) containing elements from the symbol set and asked to draw them as quickly as possible. In another testing scenario, the operator will be given a set of hand-drawn symbols and asked to identify the entity represented. The symbology will be assessed in terms of the speed and accuracy of operator performance in the two scenarios.

IV. Schedule

It is estimated that the operator testing portion of the validation will require twelve months to execute, from the time funding is identified and the Symbology Standards Management Committee (SSMC) identifies an executive agent for performing the assessment (see assumption #3) until the results are reported back to the SSMC. The following schedule of key activities is provided:

Month 1: Identify an executive agent; prepare and approve detailed test plan; identify performance instrumentation software to be developed (see assumption #4).

Month 2-6: Complete development of instrumentation software; identify operational scenarios, create test protocols and training materials, arrange for distribution to test sites.

Month 7-10: Conduct data collection at test sites (assumption #5).

Month 11-12: Perform data analysis, prepare summary report, brief results to SSMC.

V. Budget

The cost of the software instrumentation task is estimated to be $200K. In addition, each participating organization will have to contribute one labor-month of effort towards the development of operational scenarios and test protocols (to ensure that they represent the full range of operational settings in which the symbology is expected to be used).

The cost estimate for development of testing materials, data collection, analysis, and reporting is estimated to be $600K. In addition, each participating organization will be expected to provide access to space, workstations, and operators and participate in the oversight of data collection efforts at each site during the testing period.

VI. Assumptions/Risks

1. The specific tactical hardware and software to be used in the validation testing has not yet been identified. One option would be to implement the symbology and add performance instrumentation capabilities to the mapping software in the Defense Information Infrastructure Common Operating Environment. The test suite could then be installed on any of the hardware configurations supported by the Global Command and Control System (GCCS), and each participating organization could make use existing hardware to perform testing at any of its facilities where GCCS-based systems are installed. Another option would be to implement the symbology in one or more simulators in order to capitalize on performance measurement capabilities already resident in these systems; in this case, testing would be conducted at facilities where these simulators are currently available.

2. The current test plan assumes that training on the new symbology can be automated along with the rest of the data collection. The effectiveness of this approach in producing an acceptable level of proficiency with the new symbology will need to be determined. If automated training is found to be ineffective, it is possible that a data collection coordinator will be needed at each test site to deliver the symbol training and ensure the testing is conducted as planned. This form of oversight will increase the cost (in labor and travel) of the assessment and likely require an extension in the length of the data collection period.

3. The SSMC will identify an executive agent who has overall responsibility for performing the validation. The agent will prepare a detailed test plan, coordinate the development of instrumentation software and testing materials, provide oversight during testing, and produce the final test report. Members of the SSMC will approve the test plan prior to its implementation and receive periodic reports on test progress.

4. It is not known at this time whether access to source code will be required in order to develop the performance measurement module. If access is required, it may be necessary to award the task to the developer who "owns" that software, with the ability to complete the task according to the proposed schedule contingent upon developer availability. If access to source code is not required, the choice of developer will be less constrained; however, more time may be required to complete the task because the developer is unfamiliar with the software (resulting in possible slippage in the schedule).

5. Each participating organization will select the operators and locations for its portion of the validation testing. Locations are expected to include laboratory environments, command centers, simulation facilities, and field sites; operators may be specifically selected to participate in the data collection or may be included as part of a larger exercise (e.g., Prairie Warrior). The executive agent for the validation will monitor activities at the test sites to ensure that data collection is being conducted in accordance with the test plan.